



King's Research Portal

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Wongkoblap, A., Vadillo Nistal, M. A., & Curcin, V. (Accepted/In press). Predicting Social Network Users with Depression from Simulated Temporal Data. In *Proceedings - 18th International Conference on Smart Technologies, IEEE EUROCON 2019*

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Predicting Social Network Users with Depression from Simulated Temporal Data

Akkapon Wongkoblap¹

¹*Department of Informatics*

Kings College London

London, UK

akkapon.wongkoblap@kcl.ac.uk

Miguel A. Vadillo^{2,3}

²*Departamento de Psicología Básica*

Universidad Autónoma de Madrid

Madrid, Spain

miguel.vadillo@uam.es

Vasa Curcin^{1,3}

³*School of Population Health and*

Environmental Sciences

Kings College London

London, UK

vasa.curcin@kcl.ac.uk

Abstract—Mental health issues are widely accepted as one of the most prominent health challenges in the world, with over 300 million people currently suffering from depression alone. With massive volumes of user-generated data on social networking platforms, researchers are increasingly using machine learning to determine whether this content can be used to detect mental health problems in users. This study aims to investigate whether training a predictive model with multiple instance learning (MIL) via Long Short-Term Memory (LSTM) and gated recurrent unit (GRU) can improve the performance of a predictive model to detect social network users with depression. The power of MIL is to learn from user-level labels to identify post-level labels. By combining every possibility of posts label category, it can generate temporal posting profiles which can then be used to classify users with depression. This study highlights that training a MIL model via LSTM and GRU can improve the accuracy of a MIL model trained with convolutional neural networks.

Index Terms—Depression, Mental Health, Social Network, Facebook, Machine Learning, Predictive Model

I. INTRODUCTION

According to World Health Organization (WHO), the total number of people with depression globally was estimated to exceed 300 million in 2015 [1]. Only in the United Kingdom, 16% of residents experience depression at some point in their lives [2]. Mental health problems are also predicted to cost 16.3 trillion USD between 2011 and 2030, through services, treatments, and a decline in productivity at work [3].

Diagnosis of mental disorder is a challenging task which can only be done by health professionals. For their disorder to be correctly ascertained, patients need to recall how they felt and what happened to them in the previous time period, which helps the clinicians obtain comprehensive background information. However, this method is time-consuming and error-prone as sometimes patients may not be able to correctly recall their experiences. An alternative way of collecting data on symptoms of mental illness is using a self-report questionnaire, whereby people can use the questions to obtain the data themselves, either as a one-off task or regularly.

Text classification is a task of classifying any given text or message into its relevant categories. Early work of text classification has been focused on classified long documents

into categories. With its successful performance, text classification models have been applied in many areas including spam filtering, fraud detection, and news categories. In healthcare settings, huge amounts of textual data e.g., electronic health records increasingly generated are useful for researchers to develop text classifiers to identify information related to disease. Text classification models have been developed to label created content with categories, with some labelling social media posts related to users mental health [4, 5].

With massive volumes of generated data being produced on social networking platforms, researchers are increasingly studying how these data relate to user behaviour [6, 7]. By focusing on user-generated messages, it is possible to screen for users with depression through social network data [8, 9, 10]. This alternative method of detecting depressed users from their produced content can supplement the traditional diagnostics based on recollection data, by offering deeper insights into users past activities, behaviours, and feelings.

However, researching in detecting online users with depression usually has labels of each user, but lacks labels of each post. This brings the interest in developing a predictive model which can label both user-level labels and post-level labels.

Multiple instance learning (MIL) is one of possible machine learning models able to perform the task. The basic idea of MIL is to learn from a set of labelled bags, so the training does not require the individual labels in the training instances, but only labelled bags of the training set, which sets it apart from supervised learning techniques that need to know the labels of all instances [11]. MIL has been shown to successfully produce predictive models to classify sentiments on online review posts and identify sentiments on the sentence level by using only review-level labels [12, 13]. MIL has been successfully applied to detect users with depression [14]. The MIL paradigm is suitable for our dataset, since it only had labels for the users but not for their individual posts.

Recurrent neural network (RNN) is a deep learning algorithm able to learn sequential data. RNN was modified and reinvented as Long Short-Term Memory (LSTM) and gated recurrent unit (GRU) to improve its performance able to learn longer sequences. These techniques have been successfully applied in text classification [15, 16], and image caption generation [17].

This paper aims to investigate whether training a predictive model with MIL via LSTM and GRU can improve the performance of a predictive model to detect social network users with depression. The key difference to the previous work [14] is that this study uses LSTM instead of convolutional neural networks (CNNs) in middle layers to develop a MIL predictive model.

II. METHODS

This section describes the dataset used in this study, the method to measure symptoms of depression and label our training set, the architecture of our model, and the experimental setup.

A. Dataset

This study used social network data from Facebook users to build a predictive model for detecting depression symptoms. The dataset was taken from the myPersonality project, obtained from participants who took a series of psychometric questionnaires, including the Center for Epidemiological Studies Depression (CES-D), and gave consent for this data to be shared. Some of them also gave permission for their Facebook profile data to be included. Their published content was collected from 2007 until 2012 [18]. The dataset received institutional review board approval [19].

The dataset contains 6,561 submissions of CES-D from 5,947 unique participants. Removing participants who withheld permission for their Facebook profiles to be included, 939 users remained in our dataset. To ensure that there are enough patterns to distinguish users between the two groups, users who published fewer than 100 posts on their timelines were excluded, leaving the total of 509 users in the final dataset.

B. Depression Symptom Measure

The CES-D questionnaire is one of standardised and popular tools to measure depressive symptoms of respondents who take it. It comprises 20 multiple answer questions, each of them asking respondents to rate how often they experienced certain symptoms over the past week, e.g. I felt that I could not shake off the blues even with help from my family or friends.; I felt that everything I did was an effort.; I had crying spells.; I felt hopeful about the future.. After every four items the wording of questions is reversed between positive and negative phrasings. Each answer has a score between 0 and 3 e.g., 0 = Rarely or none of the time (less than 1 day), 1 = Some or a little of the time (1-2 days), 2 = Occasionally or a moderate amount of time (3-4 days), and 3 = Most or all of the time (5-7 days). The total scores can then range between 0 and 60 [20]. Respondents with scores above the cut-off (typically between 16 and 24) are then classified as depressed. This study used the cut-off score at 22. This resulted in receiving 163 users without depression and 346 users with depression.

C. Predictive Model

The predictive model was completely trained using multiple instance learning (MIL) neural networks without manual feature engineering.

The architecture of our proposed model is inspired by and follows the hierarchical attention network (HAN) introduced by [21] and the multiple instance learning network (MILNET) proposed by [12, 13]. The models have been shown to successfully perform sentiment analysis of online reviews. The concept of MILNET and its application can thus also be useful for developing a depression classifier. MILNET learns to analyse sentiment in a document from its encoding sentences or segments and then represents those as a document vector. Additionally, the model can identify the sentiment polarity of each segment of a given document. We adapt the MILNET approach by replacing segments with posted messages and a document vector with a user representation. Our proposed architecture consists of post encoder, post classification, user encoder, attention mechanism, and user classification (see Fig. 1).

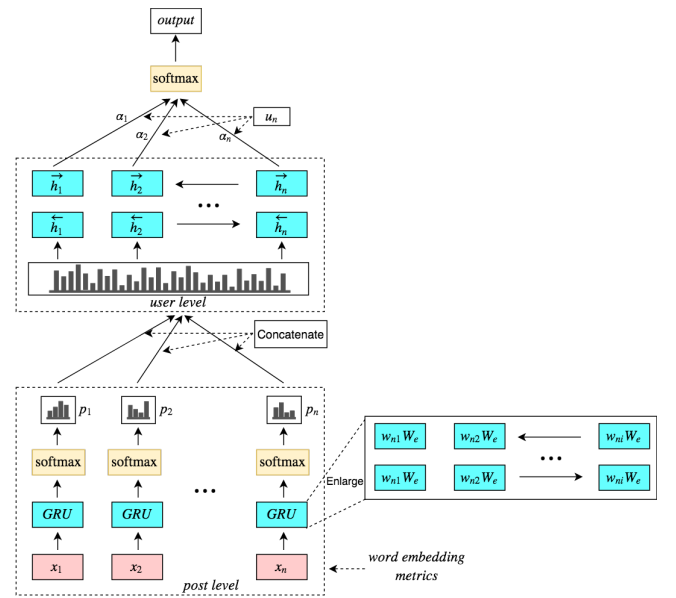


Fig. 1. The architecture of the multiple instance learning model with GRU for detecting users with depression. A posted message in the n -th post x_n was classified through GRU to p_n , which was concatenated to h_n as user representation. It was then produced as the importance of the n -th post with the context vector of post level u_n .

1) *Post Encoder*: The first layer of our model transforms raw user post data into machine readable form. Word embedding was used to transform posts to word embedding matrices. A user publishes n posts, $p_1 \dots p_n$ and each post contains i words. w_{ni} represents the word i in the n -th post. w_{ni} embedded through an embedding matrix W_e received x_{ni} . This layer embeds all words w_{ni} of n -th post to vectors:

$$x_n = w_{n1}W_e \dots w_{ni}W_e$$

After receiving word embedding vectors, a bidirectional GRU is used to encode the vectors:

$$\vec{V}_n = \overrightarrow{GRU}(x_n)$$

$$\overleftarrow{V}_n = \overleftarrow{GRU}(x_n)$$

Passing through the GRU results in post representation V_n concatenated from \vec{V}_n and \overleftarrow{V}_n . The post encoding is then sent to the post classification part to perform sentiment analysis.

2) *Post Classification*: After obtaining post representation, each post is classified based on whether it is mental health-related or related to another topic. To perform the classification, a softmax function [22] is applied to make separate predictions for every user post.

$$p_n = \text{softmax}(W_c V_n + b_c)$$

The function generates post classification $p_n = p_1^C \cdots p_n^C$, where $C \in [0, 1]$ represents the sentiments with 1 denoting a mental health related post and 0 representing a non-mental-health related topic. The parameters W_c and b_c are learnt and updated during the training step. After identifying individual post sentiments, every identified post label can be concatenated to generate a series of possibilities of post type.

3) *User Encoder*: The series of post label predictions, called user representation in this study, is encoded to summarise the changing patterns of text categories over observation time. The user representation is received by concatenating all post label possibilities of a user. A bidirectional GRU is applied through the forward hidden state and the backward hidden state:

$$\vec{h}_n = \overrightarrow{GRU}(p_n)$$

$$\overleftarrow{h}_n = \overleftarrow{GRU}(p_n)$$

Produced vectors \vec{h}_n and \overleftarrow{h}_n , are then concatenated to $h_n = [\vec{h}_n, \overleftarrow{h}_n]$.

4) *Attention Mechanism*: However, not all posts of a user convey a user characteristic. Some posts may contain cues that can be relevant to depression while others may not. For that purpose, we require the attention mechanism to be applied to reward posts that correctly represent the characteristic and are important to correctly detect a user with depression. A one-layer multilayer perceptron (MLP) produces as an attention vector of the n -th posts.

$$u_n = \tanh(W_w h_n + b_w)$$

$$\alpha_n = \frac{\exp(u_n^\top u_a)}{\sum_t \exp(u_t^\top u_a)}$$

The importance of a post (α_n) is measured as the similarity of u_n with the context vector of post level u_a , which is learnt and updated during the training step.

5) *User Classification*: Finally, a user vector can be achieved through summarising all the information of post label possibilities of a user. The user vector v is computed as follows:

$$v = \sum_t \alpha_n p_n$$

where α_n denotes the importance weight of a post and p_n represents the prediction of a type of the post. This results in obtaining a classifier to detect users with depression.

D. Experimental Setup

The baseline model was trained using the Keras library, the Python library for neural network APIs, with Tensorflow backend [23, 24]. The word embedding dimensionality was set to 25. The word embedding matrices were weighted and received from pre-trained word vectors of Glove which was trained from 2 billion tweets with 1.2 million vocab [25]. Adaptive moment estimation or Adam [26] and categorical cross-entropy were leveraged to minimise loss. All text was converted to lowercase. Numbers and URL-s in posts were replaced with the text numbers and url, respectively. Every post of users in our dataset was tokenised, and the word length per post was limited to 50 tokens. The number of posts from every user was set to 1000 to train the model. Users with fewer than 1000 posts were padded user matrices with the same length, filled with 0 values. Finally, the model was fitted with cross-validation to train and test. Due to the highly imbalanced dataset between users with and without depression, class weights were computed to weight the model during the training state. To avoid overfitting, early stopping and dropout were applied.

III. RESULTS

This section describes the results of our predictive model. To report the performance of the model, n-fold cross validation is used. Results of evaluation are presented with accuracy, precision, recall, and f1-score achieved by the model after training and testing with 5-fold cross validation.

Table I presents the results of our predictive model. Two of five folds can achieve the maximum accuracy at 75.49%, while the model achieves the average accuracy of 74.65%. The model achieves the highest results of precision of 76%, recall of 75%, and F1 score of 74%. The average results of precision, recall, and F1 score are 74%, 74%, and 72%, respectively. It is noted that our dataset is highly imbalanced. Therefore, the baseline assessment would yield approximate accuracy of 68%, in the case of the model predicting only the majority class.

TABLE I
CLASSIFICATION METRICS OF OUR PROPOSED MODEL AND THE BASELINE MODEL

	Accuracy	Precision	Recall	F1
Fold 1	74.76%	0.74	0.75	0.72
Fold 2	75.49%	0.76	0.75	0.72
Fold 3	75.49%	0.74	0.75	0.74
Fold 5	74.26%	0.73	0.74	0.71
Fold 5	73.27%	0.72	0.73	0.70
Average	74.65%	0.74	0.74	0.72
Baseline model [14]	70.54%	0.68	0.71	0.62

Fig. 2 presents ROC curves of evaluating the proposed model. The model achieves average AUC of 75% with the standard deviation of 0.03. Our model yields the highest AUC of 79%. This highlights that our baseline model can perform better than chance. It can see that the results from the model

present all the ROC curves above the red line or a random guess line.

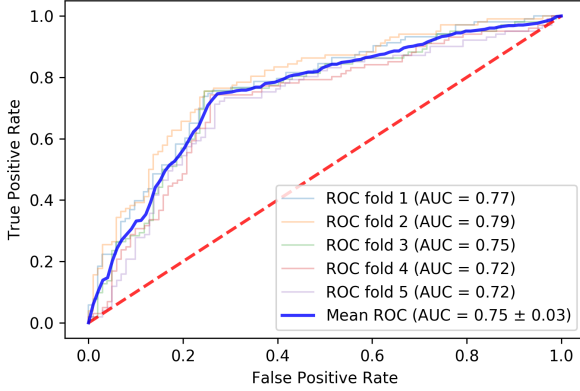


Fig. 2. ROC curves of every testing fold of our proposed model

IV. DISCUSSION

The purpose of this study was to develop a MIL predictive model with LSTM layers to detect social network users with depression. This study replaced the CNN post encoder layer of the baseline model [14] with LSTM.

We trained our proposed model and compared results with the baseline model. From Table I, it can see that our proposed model with LSTM layers can improve the accuracy from 70.54% to 74.65%. Our proposed model also achieves the higher precision to 6%, the increasing recall to 3%, and the better F1-score to 10% from the baseline model. This shows the improvement in all dimensions.

After training the proposed model, we extracted the user encoder layer which concatenates every classified post-label received from the post classification layer. The purpose of the post classification layer is to label every given post of a user into either general or mental health related text. The possibility of being general or mental health related text are estimated from given user labels. This can generate user representation. We explored the user encoder layer whether our model could generate different patterns in users with and without depression.

Fig. 3 shows changing patterns of publishing posts related to general or mental health topics from the proposed model. The figure represents the patterns from 6 random users with depression and non-depression. The y-axis of the chart shows the possibilities of being another topic denoted with 0 and mental health related text denoted with 1, and the x-axis is the number of posts. These patterns are generated from users who were correctly labelled by our model. On the left-hand side, it can see that users with non-depression tended to have fluctuating changes over time. Considering users with depression, the changing patterns were more stable and most of their posts labelled with mental health related text. This highlights that the model could use the patterns to distinguish between the two groups of users.

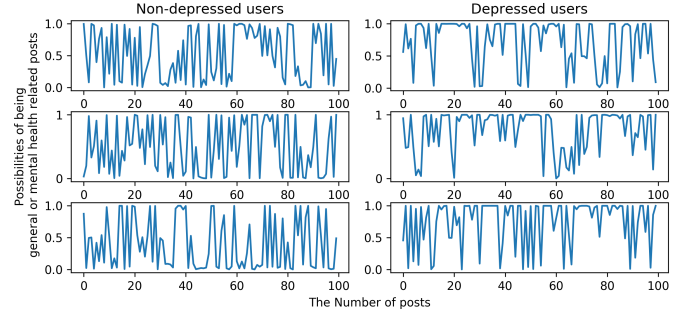


Fig. 3. The patterns of general and mental-health related posts in 3 depressed and 3 non-depressed users from our proposed model. The y-axis of the chart shows the possibilities of being another topic denoted with 0 and mental health related text denoted with 1. The x-axis shows the number of posts.

Comparing parameters of our proposed model with the baseline model [14], the latter work used 500 posts from every user to train their model, while our work used 1000 posts. Another parameter was the dimensions of pre-trained word vectors. This study used 25 dimensions trained from 1.2 million vocab of tweets, while the previous work used 100 dimensions trained from 400K vocab of Wiki and English Gigaword, a comprehensive archive of newswire text data. This may be why our model could achieve better performance.

This study has some limitations. Our dataset was highly unbalanced, as it had a higher number of depressed users than what is found in general population. Another limitation is that the model was trained with a relatively few of users, and the performance may be improved if it is trained on a larger sample. Our last limitation is that our dataset does not reflect real world population of patients with depression.

Future work is planned to develop our proposed model via transfer learning which is the method of reuse of a well-trained machine learning model in one task on another related task. Applying transfer learning is useful when a starting task has enough training data but limited for another related task. We are also planning to collect more data to explore whether our proposed model is generalised when is applied to other datasets.

V. CONCLUSION

In this paper, we developed a MIL predictive model to detect social network users with depression from LSTMs. It found that our proposed model achieved the average accuracy of 74.65% in detecting depressed users from their social network created content with additionally generating changing patterns over observation time. We can improve our proposed model compared to a MIL model with CNNs in all dimension of precision, recall, f1-score, and average accuracy. To our best knowledge, this study is the first one to apply the MIL model with LSTMs and generate temporal data from created text to detect social network users with depression.

The model could potentially be applied to structured and unstructured text and map it to longitudinal data, which can provide better understandings of changing patterns of text over

observation time and have a considerable impact on health care research, e.g., using free text from electronic health records (EHRs) to generate longitudinal phenotypes for research.

REFERENCES

- [1] World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," 2017.
- [2] Mental Health Foundation, *Fundamental Facts About Mental Health 2016*. London: Mental Health Foundation, 2016.
- [3] D. E. Bloom, E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, R. L. Bloom, S. Fathima, B. A. Feigl, T. Gaziano, M. Mowafi, A. Pandya, K. Prettnner, L. Rosenberg, B. Seligman, A. Z. Stein, and C. Weinstein, "The Global Economic Burden of Noncommunicable Diseases," pp. 1–46, 2011.
- [4] G. Gkotsis, A. Oellrich, S. Velupillai, M. Liakata, T. J. P. Hubbard, R. J. B. Dobson, and R. Dutta, "Characterisation of mental health conditions in social media using Informed Deep Learning," *Scientific Reports*, vol. 7, p. 45141, mar 2017. [Online]. Available: <http://www.nature.com/articles/srep45141>
- [5] J. Ive, G. Gkotsis, R. Dutta, R. Stewart, and S. Velupillai, "Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. Association for Computational Linguistics, 2018, pp. 69–77. [Online]. Available: <http://aclweb.org/anthology/W18-0607>
- [6] Long Jin, Yang Chen, Tianyi Wang, Pan Hui, and A. V. Vasilakos, "Understanding user behavior in online social networks: a survey," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 144–150, sep 2013. [Online]. Available: <http://ieeexplore.ieee.org/document/6588663/>
- [7] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. E. Zhao, "Analyzing patterns of user content generation in online social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*. New York, New York, USA: ACM Press, 2009, p. 369. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1557019.1557064>
- [8] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 51–60. [Online]. Available: <http://www.aclweb.org/anthology/W14/W14-3207> <http://aclweb.org/anthology/W14-3207>
- [9] M. De Choudhury and M. Gamon, "Predicting Depression via Social Media," *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, vol. 2, pp. 128–137, 2013.
- [10] A. Wongkoblaph, M. A. Vadillo, and V. Curcin, "Researching Mental Health Disorders in the Era of Social Media: Systematic Review," *Journal of medical Internet research*, vol. 19, no. 6, p. e228, jun 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28663166>
- [11] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, may 2018.
- [12] S. Angelidis and M. Lapata, "Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis," *Transactions of the Association for Computational Linguistics*, vol. 6, pp. 17–31, 2018.
- [13] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, "From Group to Individual Labels Using Deep Features," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. New York, New York, USA: ACM Press, 2015, pp. 597–606. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2783258.2783380>
- [14] A. Wongkoblaph, M. A. Vadillo, and V. Curcin, "Modeling depression symptoms from social network data through multiple instance learning," *AMIA Summits on Translational Science Proceedings*, in press.
- [15] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Advances in neural information processing systems*, 2015, pp. 649–657.
- [16] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," *arXiv preprint arXiv:1605.05101*, 2016.
- [17] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3156–3164.
- [18] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *The American psychologist*, vol. 70, no. 6, pp. 543–56, sep 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26348336>
- [19] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1036–1040, jan 2015. [Online]. Available: <http://www.pnas.org/lookup/doi/10.1073/pnas.1418680112>
- [20] L. S. Radloff, "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement*, vol. 1, no. 3, pp. 385–401, jun 1977.
- [21] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of*

the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1480–1489.

- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.
- [23] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “Tensorflow: A system for large-scale machine learning,” in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, ser. OSDI’16. Berkeley, CA, USA: USENIX Association, 2016, pp. 265–283. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3026877.3026899>
- [24] F. Chollet and Others, “Keras,” 2015.
- [25] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation,” in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: <http://www.aclweb.org/anthology/D14-1162>
- [26] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *CoRR*, vol. abs/1412.6, 2014. [Online]. Available: <http://arxiv.org/abs/1412.6980>